Analysing User Lifetime in Voluntary Online Collaboration

McHugh, Ronan; Larsen, Birger

Published in:
Proceedings of the Tenth Danish Human-Computer Interaction Research Symposium (DHRS2010)

Publication date:
2010

Document version
Early version, also known as pre-print

Analysing User Lifetime in Voluntary Online Collaboration

Ronan McHugh
Royal School of Library and Information Science
Birketinget 6, 2300 Copenhagen S.
Denmark
mchugh.r@gmail.com

Birger Larsen
Royal School of Library and Information Science
Birketinget 6, 2300 Copenhagen S.
Denmark
blar@iva.dk

ABSTRACT
This paper analyses persuasion in online collaboration projects. It introduces a set of heuristics that can be applied to such projects and combines these with a quantitative analysis of user activity over time. Two example sites are studies, Open Street Map and The Pirate Bay. Results show that there is a correspondance between some features of site design and user participation patterns in the projects.

Keywords
persuasive design, online collaboration, heuristics.

1. INTRODUCTION
This paper analyses how user activity changes over time in two collaborative online projects, The Pirate Bay and Open Street Map. The paper combines two analyses: first, a heuristic analysis of site design assesses persuasive features of the projects, second, a quantitative analysis of user histories looks at how user participation patterns varies between the two projects.

1.1 Online Collaboration
In this paper, online collaboration refers to the phenomenon of voluntary participation in cooperative projects co-ordinated over the Internet. The phenomenon is represented in projects such as Wikipedia[1], a collaboratively written encyclopedia and Peer-to-Peer University[2], an online education resource. In general, such projects have several common characteristics, namely, that contributions are made on a voluntary basis, coordination of work tends to be organic and without formal hierarchy, and that projects are typically based around specific values.

Online collaboration projects are a fascinating field of study and present a unique opportunity for researchers. The digital nature of participation means that an unprecedented level of data is available for analysis, as user actions are recorded by the system in question and are often freely accessible. This data thus allows us to apply quantitative analysis to the study of human creativity and innovation.

1.2 Persuasive Design
In order to analyse participation in the projects in question, we employ the concept of Persuasive Design. Persuasive Design is concerned with the ways in which computers and related devices can alter user behaviour through psychological processes. B.J. Fogg, the founder of the discipline, defines persuasion as "an attempt to change attitudes or behaviours or both"[1]. Utilising the perspective of Persuasive Design in this research means looking at the ways in which projects encourage users to participate.

Fogg proposes the Fogg Behaviour (FBM) Model for analysing persuasion[2] . According to Fogg, successful persuasion is dependent on three factors: a user's motivation to perform a target action, their ability to do so and triggers to perform the action made by the system. Designing persuasive technologies involves boosting motivation or ability or both, while also ensuring that the desired behaviours are triggered at the appropriate time.

The clarity of this model is very useful for helping one think about persuasion in a dynamic way; one can see how site designers can remove barriers to ability or attempt to increase motivation as part of their persuasive strategy. On the other hand, the model seems most applicable to conceptualising one-off persuasive goals, such as persuading users to click on a sign-up link, or to purchase a product. The model is less intuitively useful when applied to large-scale collaborative projects, which involve repeated actions by users over an extended period of time. Since our cases do not deal with one-off actions such as buying a book or something similar, it seems appropriate to frame the model differently to account for these differences. Below, in section 2.1, we propose a set of heuristics to fit online collaboration sites.

1.3 Summary of Cases
The quantitative analysis is based on complete user histories downloaded from two online collaborative projects; Open Street Map[3] (OSM) and The Pirate Bay[4] (TPB). Open Street Map is a collaboratively produced map of the world. Participants contribute by adding points to the map which they may have derived from exploring an area with a GPS transmitter or simply from local knowledge. The Pirate Bay is a site which indexes torrent files which are used to download files collaboratively, from multiple computers at a time. Participants contribute by uploading torrent files and allowing other users to download files from their computer.

2. METHODOLOGY
This paper uses two distinct methodologies: on the one hand, a set of heuristics are employed to conduct an analytic walkthrough of participation in the sites. On the other hand, complete participation histories for a sample of users of both projects are downloaded and analysed.

2.1 Heuristic Analysis
The heuristics used in this research are based on the Fogg Behaviour Model, adapted to take account of the peculiar nature of online collaboration projects. As stated above, online collaboration projects share several characteristics, 1) contributions tend to be made on a voluntary basis, 2) work tends to be coordinated organically by participants without the use of rigid hierarchy, and 3) projects are strongly driven by specific values which are embodied in their work. These facets have

[1] http://wikipedia.org
[2] http://p2pu.org/
[3] http://www.openstreetmap.org
[4] http://thepiratebay.org

several implications for project design and thus for the heuristics needed to analyse these.

To start with, we can see that the voluntary nature of contribution requires high levels of motivation among participants. Projects thus need to develop strong motivational support for participation. Secondly, if work is to be coordinated in a decentralised way, then projects need to facilitate a high level of communication and coordination between participants. The value driven nature of projects can serve as a motivational tool for increasing participant motivation, but these values need to be reflected by the site design and also in the products of the projects in question. Based on these aspects and on the FBM, we suggest the following heuristics for analysing design of collaborative online projects.

1. The values that underpin the site should be clearly visible to all users and should be reinforced regularly.

2. Triggers to participate in the project should be visible to users of the product. Triggers should recur throughout user lifetime and not just be visible to entry-level users.

3. Participation in the project should be as simple as possible and documentation of technical aspects should be easily available.

4. The project should encourage users to identify themselves with the project and feel as if their contributions are valued through rewards, feedback or other mechanisms.

5. The interface should facilitate interaction between users and coordination of collaborative efforts.

These heuristics are applied to our cases in section 3.1 below.

2.2 Quantitative Analysis

2.2.1 Data Retrieval

The data for this study was retrieved by downloading histories of user activities stored publicly on the websites in question. URLs for user profiles were obtained by entering the unique sub-directories for user profiles into Yahoo! SiteExplorer[footnote: http://siteexplorer.search.yahoo.com/] and downloading the first 1,000 results, which are the maximum that can be downloaded. [footnote: Since TPB stores user profiles at two different subdomains, it was possible to download a larger sample.] Duplicates were removed and a script based on Python's 'Beautiful Soup' module was used to download the full histories associated with each user, converting pages from a html format into a tabbed text file.[5]

2.2.2 Bin division of participants

In order to facilitate analysis, it was decided to divide each sample of users into three bins based on total activity levels. After analysis of the Lorenz distribution of participation rates for both projects, it was decided to divide the samples based on the formula of 60, 30, 10. The first 60% of participants are the lowest level contributors, the next 30% are medium level contributors and the final 10% are the highest level contributors. This method was chosen because of the high rate of participation inequality observed within the samples, whereby a small number of participants are responsible for a large percentage of contributions while a majority of participants only ever contribute a relatively small amount [3, 4].

2.2.3 Data Analysis

In order to analyse user participation rates over time, a series of spreadsheet formulae were used to number all user participation events according to when in the user's lifetime they occurred. Thus, all user activity could be charted on a timeline starting with their first ever contribution to the project. Using these timelines, a series of frequency distributions were derived which plotted the percentage of total contributions for each user group that occurred within a specific time-frame (e.g. two weeks, three months, etc). This methodology allowed us to make broad observations about the average lifetime participation rates of particular groups of contributors and compare these with other users of the same project or with the respective user division of the other project.

3. RESULTS

3.1 Heuristic Walkthrough

Using the heuristics outlined above, we conducted a walkthrough of participation in both projects in order to assess the persuasive strength of the sites.

1. The values that underpin the site should be clearly visible to all users and should be reinforced regularly.

– OSM - values are not very visible to surface level users but they are present in a regularly updated blog, and are often embodied in the various Projects of the Week, which set regular challenges for OSM mappers.

– TPB - prominent logo neatly captures many of values. However, there seems to be little readily accessible content discussing the broader context of file-sharing and copyright laws.

2. Triggers to participate in the project should be visible to users of the product. Triggers should recur throughout user lifetime and not just be visible to entry-level users.

– OSM - “Sign Up” and “Edit” triggers are visible from front page. The Project of the Week acts as a recurring trigger for participants.

– TPB - “Register” button visible from front page but not immediately obvious. No obvious encouragement to upload torrents present.

3. Participation in the project should be as simple as possible and documentation of technical aspects should be easily available.

– OSM - plentiful documentation for new beginners linked to in registry email, including a beginner's guide and screencast videos.

– TPB - a large selection of tutorials are provided in the forum, where it is also possible to ask further questions.

4. The project should encourage users to identify themselves with the project and feel as if their contributions are valued through rewards, feedback or other mechanisms.

[5] Beautiful soup is a Python module specifically designed for screen-scraping HTML pages http://www.crummy.com/software/BeautifulSoup/

– OSM - Symbolic rewards are given to users based on the number of points they have uploaded, however these are on user pages and are not particularly obvious.

– TPB - Users can achieve Trusted and VIP status based on their contributions to the site. In the forums users are given ratings based on how helpful their posts are.

5. The interface should facilitate interaction with other users and coordination of collaborative efforts.

– OSM - Project encourages co-ordination via OSM Wiki. However, users must search for the appropriate forum or mailing list. There is a strong focus on making connections with other OSM users in one's geographic area.

– TPB - There is a single forum which is easy to find and has a large amount of material. The forum also makes it possible for users to request torrents. The comment feature on torrents enables users to request seeders, provide links to subtitles, and rate torrent quality, among other things.

3.2 Quantitative Analysis

3.2.1 Summary of data

The TPB dataset consisted of 268,141 torrents produced by 1,495 users. The set had an average contribution of 179.36 torrents per user with a median of 10. The OSM dataset consisted of 1,884,104 edits contributed by 762 users. This gives an average of 2472.58 edits per user, with a median of 299. The average lifetime of TPB users is 308.35 days and the median is 169 days compared to 514.88 days and 516 days for OSM users.

Due to problems with the data retrieved we have only analysed low and mid-level contributors. This is due to obvious flaws in the data retrieved for the highest level contributors to OSM, whereby these contributors had improbably low lifetimes, for example, some users with many thousands of uploads had lifetimes of only eleven days. This suggests that the data retrieved was only a partial representation of their total lifetime and as such lifetime based analysis of their contributions was thought not to be representative.

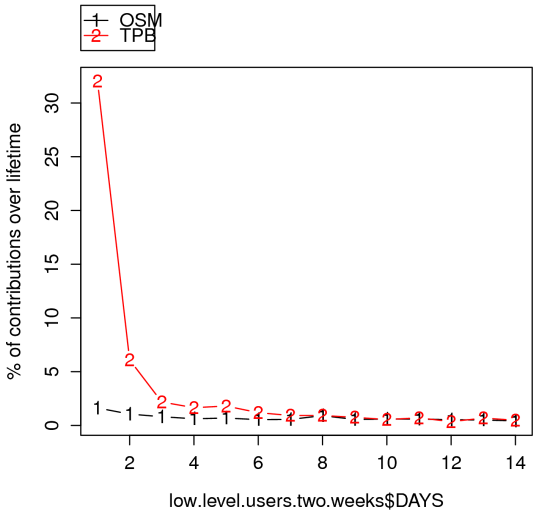

Fig 1. Low level user activity over first two weeks of lifetime

3.2.2 Comparison of low level users

Time-based analysis of contribution rates of low-level users across systems show a considerable amount of difference between the two projects. As can be seen in Fig. 1, Low-level TPB users contribute proportionally far more in the first days of their lifespans than corresponding OSM users. This difference is particularly apparent in the first two weeks of lifetime and the first day especially, where TPB users contribute 32.04% of their total uploads, while OSM users contribute only 1.62% of their total edits.

It is only after about 14 weeks that OSM contribution rates start to be significantly higher than TPB rates, with OSM editors contributing 1.28% of total lifetime edits while TPB users contribute 0.68%. This difference becomes more pronounced as time goes on, as can be seen in Fig. 2. In the period between 330 days and 360 days after first activity, OSM editors contributed 3.69% of total lifetime edits while TPB users contributed 0.88%.

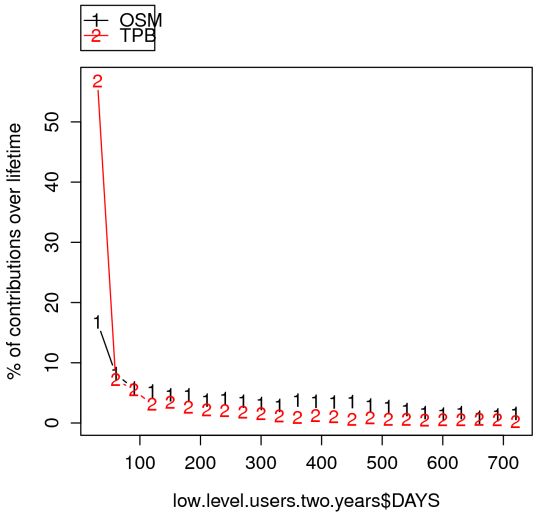

Fig 2. Low-level users activity over first two years of lifetime

This comparison points to a different dynamic of participation which can also be seen in the different lifespans of users; the median lifetime of low-level TPB users is 19 days, while the median lifetime of low-level OSM users is 432 days. 36% of low-level TPB users contribute for only one day, while only 13% of low-level OSM users do the same. These figures suggest that OSM is far better at persuading users to maintain their involvement in the project. The fact that the median lifespan of low-level OSM editors is well over a year suggests a far more sustainable level of involvement among OSM editors.

3.2.3 Comparison of mid-level users across systems

The lifespan analyses of mid-level users reveals some surprising results. As with the analysis of low-level contributors, mid-level TPB users start their activity periods by contributing more than their OSM counterparts, although the difference is not so great, 3.86% of total contributions in their first day vs 1.39% of OSM mid-level contributions (Fig. 3).

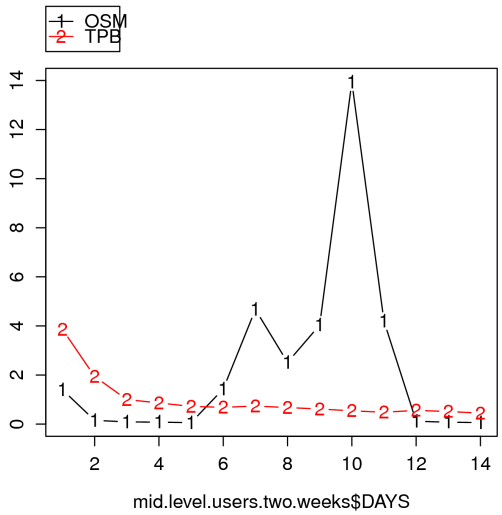

Fig 3. Mid level user activity over first two weeks of lifetime

What is surprising is the extremely large rise in OSM contributions relative to those of TPB users after the sixth day. This increase in contributions is reflected in the two year timeline (Fig. 4) where the OSM contributions are more concentrated in the early days of lifespan than those of TPB users. This huge concentration of productivity in the second week of OSM user activity leads to consistently lower productivity over the following months of activity, until 390 days where the OSM users again begin to outperform their TPB counterparts. The average lifespan of mid-level TPB users is 476.22 days, while the median is 406.5, OSM mid-level users on the other hand have an average lifespan of 784.25 days and a median of 785 days.

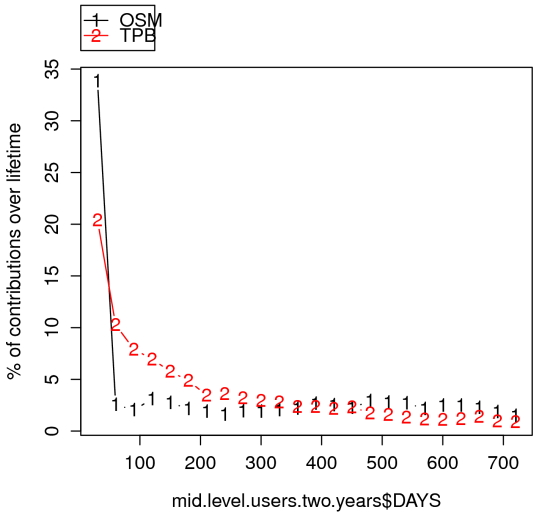

Fig. 4: Mid-level user activity over the first two years of lifetime

This indicates that despite the flurry of activity in the first week, mid-level OSM users are both longer-lasting and more consistent than their TPB counterparts.

4. CONCLUSIONS AND FUTURE WORK

The foregoing analyses have produced some interesting results. The heuristic analysis found that OSM is better than The Pirate Bay at presenting the values behind the project and reminding users of these on a regular basis. Moreover, its 'Project of the Week' feature represents a recurring trigger which constantly encourages users to maintain their activity in the project. It is also better at facilitating social gatherings and coordination between participants which may provide other forms of triggers through 'Mapping Parties' and normative pressure to contribute.

Meanwhile the quantitative analysis has shown that OSM users in general contribute over a longer period of time and at a more consistent rate than their counterparts in TPB. This result suggests that there could be a connection between OSM's better motivational and triggering features and its users' longer lifetimes. Testing this connection in a controlled setting could be a challenge for future research.

5. REFERENCES

[1] BJ Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann, 2002.

[2] BJ Fogg. A behavior model for persuasive design. In Persuasive '09: *Proceedings of the 4th International Conference on Persuasive Technology,* pages 1-7, New York, NY, USA, 2009. ACM.

[3] Jakob Nielsen. Participation Inequality: Encouraging more users to contribute http://www.useit.com/alertbox/participation_inequality. April 2009.

[4] Clay Shirky. *Here Comes Everybody: How change happens when people come together*. Penguin Books, 2008.